Genomics 114 (2022) 110500

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Comparative analysis of capture methods for genomic profiling of circulating tumor cells in colorectal cancer

Joao M. Alves [a,b,*], Nuria Estévez-Gómez [a,b], Monica Valecha [a,b], Sonia Prado-López [a,b,1], Laura Tomás [a,b], Pilar Alvariño [a,b], Roberto Piñeiro [c,d], Laura Muinelo-Romay [d,e], Patricia Mondelo-Macía [e], Mercedes Salgado [f], Agueda Iglesias-Gómez [g], Laura Codesido-Prada [g], Joaquin Cubiella [g], David Posada [a,b,h,*]

[a] CINBIO, Universidade de Vigo, 36310 Vigo, Spain
[b] Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, Spain
[c] Roche-Chus Joint Unit, Translational Medical Oncology Group, Oncomet, Health Research Institute of Santiago de Compostela (IDIS), Santiago de Compostela, Spain
[d] Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain
[e] Liquid Biopsy Analysis Unit, Translational Medical Oncology Group, Health Research Institute of Santiago de Compostela (IDIS), Santiago de Compostela, Spain
[f] Department of Oncology, Hospital Universitario de Ourense, Research Group in Gastrointestinal Oncology-Ourense, Ourense, Spain
[g] Department of Gastroenterology Hospital Universitario de Ourense, Research Group in Gastrointestinal Oncology-Ourense, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Ourense, Spain
[h] Department of Biochemistry, Genetics, and Immunology, Universidade de Vigo, 36310 Vigo, Spain

ARTICLE INFO

Keywords:
Circulating tumor cells
Liquid biopsy
Intratumor genomic heterogeneity
Colorectal cancer
Precision medicine

ABSTRACT

The genomic profiling of circulating tumor cells (CTCs) in the bloodstream should provide clinically relevant information on therapeutic efficacy and help predict cancer survival. Here, we contrasted the genomic profiles of CTC pools recovered from metastatic colorectal cancer (mCRC) patients using different enrichment strategies (CellSearch, Parsortix, and FACS). Mutations inferred in the CTC pools differed depending on the enrichment strategy and, in all cases, represented a subset of the mutations detected in the matched primary tumor samples. However, the CTC pools from Parsortix, and in part, CellSearch, showed diversity estimates, mutational signatures, and drug-suitability scores remarkably close to those found in matching primary tumor samples. In addition, FACS CTC pools were enriched in apparent sequencing artifacts, leading to much higher genomic diversity estimates. Our results highlight the utility of CTCs to assess the genomic heterogeneity of individual tumors and help clinicians prioritize drugs in mCRC.

1. Introduction

Although research on cancer biology has traditionally been hampered by sampling issues, with most approaches relying on highly invasive, risky, and, sometimes, difficult to obtain solid tissue biopsies [1–3], strong evidence has emerged in recent years that the peripheral blood, as well as other body fluids, offer a valuable source of cancer-associated materials [4,5]. As opposed to tissue biopsies, liquid biopsies represent a minimally invasive alternative to capture clinically-relevant information about tumors [6], including circulating tumor cells (CTCs). CTCs are thought to consist of cells shed by the primary tumor (PT) and metastatic lesions into the bloodstream and have become the subject of intense research due to their potential role in the metastatic process [6–8]. In recent years multiple studies have demonstrated the clinical significance of CTCs for prognosis and therapeutic management, with CTC burden being correlated with unfavorable overall survival in several cancer types [9–12].

Importantly, sequencing studies exploring the genomic landscape of CTCs [12]; [13–15] have shown that the mutational profiles of single CTCs generally reflect the overall genomic composition of both matched primary and metastatic lesions. In some cases, CTC genomic diversity might represent intratumoral heterogeneity better than single tumor tissue biopsies [16]. Incorporating CTC genomic information is expected to increase the clinical value of liquid biopsies by providing better

* Corresponding authors at: CINBIO, Universidade de Vigo, 36310 Vigo, Spain.
E-mail addresses: jalves@uvigo.es (J.M. Alves), dposada@uvigo.es (D. Posada).
[1] Current address: Institute of Solid state Electronics, Technische Universität Wien, Austria.

https://doi.org/10.1016/j.ygeno.2022.110500
Received 10 June 2022; Received in revised form 23 September 2022; Accepted 2 October 2022
Available online 4 October 2022

predictions of therapeutic sensitivity and survival outcomes [17,18].

Nevertheless, implementing a comprehensive molecular characterization of CTCs into routine clinical procedures has, so far, proven extremely challenging [19,20]. Indeed, despite the numerous strategies already available for isolating CTCs [21], ranging from methods based on the physical properties of cells (i.e., size and deformability) to others based on biological characteristics (e.g., cell surface marker expression), CTCs are typically present at low numbers in the blood and can show a wide range of phenotypes [20,22,23]. Consequently, these technologies are likely to differ in their detection sensitivity and recovery rates, but, to date, comparative analyses across distinct technologies are lacking [22]. Of particular interest to us, it remains unclear how the different isolation methods can impact the assessment of the genomic landscape of CTCs.

To identify an efficient strategy for the downstream genomic profiling of CTCs in metastatic colorectal cancer (mCRC), we contrasted three CTC-enrichment approaches —CellSearch®, Parsortix®, and Fluorescence-Activated Cell Sorting (FACS). While all methods evaluated here struggled with data quality issues, our results indicate that Parsortix and, in part, CellSearch can provide genomic heterogeneity scores, mutational signature profiles, and therapeutic targets compatible with those found in matching primary tumor samples.

2. Material & methods

2.1. Patient selection and blood collection

We enrolled four mCRC patients diagnosed between October 2017 and September 2019 at the Hospital Universitario de Ourense, Spain, with histologically proven CRC and either therapy-naive or showing evidence of progression. On the same day, we collected three blood samples per patient. We stored them in three different containers, one for each CTC-enrichment protocol, at room temperature: CellSave Preservative tubes (Menarini Silicon Biosystems, Italy) for CellSearch, Transfix CTC-TVT tubes (Cytomark, UK) containing formaldehyde for Parsortix and cell-free DNA BCT CE tubes (Streck, NE, USA) for FACS. In addition, we obtained a formalin-fixed paraffin-embedded (FFPE) block of the primary tumor (PT) from each patient. Importantly, all specimens were obtained and collected after written informed consent from all subjects using a protocol approved by the Clinical Ethics Committee of Pontevedra-Vigo-Ourense (2018/301 approved 19/06/2018).

2.2. CTC enrichment

We drew the blood using three different strategies for CTC enrichment. The CellSearch® system (Menarini, Silicon Biosystems, Bologna, Italy) enumerates and isolates CTCs of epithelial origin (CD45-, EpCAM+, and CK8+, 18+, and/or 19+). The Parsortix® platform (ANGLE plc, UK) traps CTCs due to their larger size and lower compressibility than blood cells. The FACS strategy separates CTCs based on custom markers; in our case EpCAM+/CD45-/CK7,8 + .

2.2.1. CellSearch

We processed 7.5 mL of whole blood for each sample in the CellTracks Autoprep system using the Circulating Tumor Cell Kit (Menarini, Silicon Biosystems, Bologna, Italy). This kit consists of ferrofluids coated with epithelial cell-specific anti-EpCAM antibodies to immunomagnetically enrich epithelial cells and a mixture of antibodies directed to cytokeratins (CKs) 8, 18, and 19 conjugated to phycoerythrin (PE); an antibody to CD45 conjugated to allophycocyanin (APC); and a nuclear dye 4′,6-diamidino-2-phenylindole (DAPI). Staining steps involve the fixation and permeabilization of the cells. Afterward, according to the manufacturer's instructions, we analyzed the processed samples with the CellTracks Analyzer II. We identified the CTCs as round or oval cells with an intact nucleus (DAPI positive), CK positive, and CD45 negative (Fig. S1). We stored the CTC-enriched samples at −80 °C after recovering the cells from the CellSearch cartridge. All samples were processed within 36 h of collection.

2.2.2. Parsortix

We loaded 7.5 mL of whole peripheral blood per sample into a Parsortix microfluidic device (Angle plc, UK). We enriched the samples in disposable Parsortix cassettes with a gap size of 6.5 μm (GEN3D6.5, Angle Inc., Guildford, UK) and at 99 mbar of pressure, according to the manufacturer's guidelines. After separation, we collected the captured cells in 200 μL of PBS and stored them at −80 °C. All samples were processed within 36 h of collection.

2.2.3. Fluorescence-activated cell sorting (FACS)

To obtain the peripheral blood mononuclear cell (PBMC) fraction –to be used as healthy controls– we took 1 mL from each blood sample and performed Ficoll-Paque gradient centrifugation. We kept the PBMCs in RNA later (Ambion, TX, USA) at −80 °C until the extraction of genomic DNA (gDNA). Then we used the remaining blood volume (7–9 mL) for CTC staining and collection. After validating our FACS protocol using spike-in experiments (see Supplementary note 1), we followed a similar approach to Miller et al. [24]. First, we lysed the red blood cells using BD Pharm Lyse lysing solution (BD Biosciences, NJ, USA), following the fabricant recommendations. When needed, we repeated the lysing step up to three times. We then resuspended the cells in phosphate-buffered saline (PBS) solution and filtered them with a 70 μm cell strainer (Falcon, NY, USA). We used the FIX & PERM™ Cell Permeabilization Kit (Invitrogen, MA, USA) and incubated the filtered cell suspensions with antibodies (BD Biosciences, NJ, USA) against the epithelial cell adhesion molecule (EpCAM; PerCP-Cy5.5, IgG1λ, clone EBA-1) and the leukocyte common antigen CD45 (FITC, IgG1κ, clone HI30) with reagent A for 25 min in the dark at room temperature for fixation. Cells were washed once with 500 μL of PBS and centrifuged at 200×*g* for 5 min. We resuspended the cell pellet in 1 ml of PBS and incubated it with an antibody against the epithelial markers cytokeratins 7 and 8 (CK7,8; PE, IgG2α/κ, clone CAM 5.2) and reagent B for 20 min in the dark at room temperature for permeabilization. We washed once again and resuspended the cells in 500 μl of PBS. Finally, we selected and collected 3 μl of PBS pools of CTCs based on an EpCAM+/CD45-/CK7,8+ phenotype (Fig. S2) using a FACSAria III (BD Biosciences, NJ, USA). We analyzed the data using the FACSDiva (BD Biosciences, NJ, USA) and FlowLogic software (Miltenyi Biotec, Germany). All samples were processed within 24 h of collection.

2.3. Whole-genome amplification of CTC-pools

Given the large collection volume (~200 μl) from both CellSearch and Parsortix, we initially performed genomic DNA (gDNA) extraction of the CTC enriched samples obtained from these platforms using the QIAamp DNA Blood Mini Kit (Qiagen, Germany) before performing whole-genome amplification (WGA) using the Ampli1 kit (Menarini Silicon Biosystems, Italy). We carried out the WGA starting with 1 μl of DNA for the CellSearch and Parsortix samples and directly in the case of the FACS samples. We worked in a laminar-flow hood to avoid contamination and used a dedicated set of pipettes and UV-irradiated plastic materials. We included positive (10 ng/μl REPLIg human control kit, Qiagen, Germany) and negative controls (DNase/RNase free water) during the amplification and used the Ampli1 QC Kit to evaluate the amplification. Samples with a positive signal for at least two PCR fragments were selected to increase the total dsDNA content using the Ampli1 ReAmp/ds kit. We then removed the kit adaptors by incubating at 37 °C for 3 h a mixture of 5 μl of NEBuffer 4 10× (New England Biolabs, MA, USA), 1 μl of *Mse*I 50 U/μl (New England Biolabs, MA, USA), 19 μl of nuclease-free water and 25 μl of dsDNA followed by a step at 65 °C for 20 min for enzyme inactivation. Finally, we purified the samples with 1.8× AMPure XP beads (Agencourt, Beckman Coulter, CA, USA), quantified the DNA yield with Qubit 3.0 fluorometer (Thermo

Fisher Scientific, MA, USA), and checked the amplicon size distribution with the D1000 ScreenTape System in a 2200 TapeStation platform (Agilent Technologies, CA, USA).

2.4. FFPE and PBMCs bulk gDNA isolation

We extracted bulk gDNA from FFPE samples using the QIAamp DNA FFPE tissue kit (Qiagen, Germany) by incubating and shaking them at 60 °C for 1 h before slicing them and adding a deparaffinization solution (DS). We performed a 56 °C incubation step for 1 h and, when needed, conducted a second addition of DS in a new tube to remove the remaining paraffin. We used the QIAamp DNA Blood Mini Kit (Qiagen, Germany) for the extraction of gDNA from PBMCs bulks and estimated DNA yield using the Qubit 3.0 fluorometer (Thermo Fisher Scientific, MA, USA) and DNA integrity with the Genomic DNA ScreenTape Assay (Agilent Technologies, CA, USA).

2.5. Whole-exome sequencing

CTC pools and bulk sequencing libraries were constructed at the Spanish National Center for Genomic Analysis (CNAG; http://www.cnag.crg.eu) with the SureSelect XT and Agilent Human Exon v5 kits (Agilent Technologies, CA, USA). In total, seven whole-genome amplified CTC-pools (two from FACS: P4 and P5, one from CellSearch: P1 and four from Parsortix: P1, P3, P4, and P5) and four FFPE bulk samples were sequenced at 100× and four PBMCs samples at 60×. All samples were run on an Illumina NovaSeq 6000 (PE100) at CNAG.

2.6. Data processing and variant calling

After trimming amplification and sequencing adapters from the raw FASTQ files, we aligned the sequencing reads from CTC pools, tumor, and healthy samples to the Genome Reference Consortium Human Build 37 (GRCh37) using the MEM algorithm in the BWA software [25]. Following a standardized best-practices pipeline [26], we filtered out reads with low mapping quality. We next performed a local realignment around indels and removed PCR duplicates. We identified somatic single nucleotide variants (SNVs) for each CTC-capture method using the multi-sample variant-calling feature implemented in MuTect2 software, taking the BAM files of the available sample types (i.e., tumor bulk + CTC-pool + healthy control). We then used FilterMutectCalls to remove calls in sequence context artifacts or contamination fractions (see Supplementary note 2 and Fig. S3). Afterward, the genotypes of variants showing a coverage depth ≥ 10, alternative allelic depth ≥ 2 and allele frequency estimates ranging from 0.05 to 0.75 were kept for downstream analysis. We merged the inferred SNV calls for all datasets and performed variant annotation using Annovar software (v.20200608) [27].

2.7. Mutational signatures

For all datasets, we ran sigProfilerExtractor [28] under default parameters to identify de novo mutational signatures for single-base substitutions (SBS), followed by the assignment of the decomposed signatures to known COSMICv3 SBS96 signatures [29].

2.8. PanDrugs

We used PanDrugs [30] (http://www.pandrugs.org) –a web-based platform that attempts to match genomic data to available drug therapies to guide personalized treatment selection– to explore changes in therapeutic options and drug suitability scores across the different datasets. For that purpose, we first ran PanDrugs using a VCF with the list of exonic mutations identified in the tumor bulk sample of each patient to identify CRC-specific therapeutic candidates. Next, we extracted the drug score (which ranges from −1 to 1 and measures the suitability of each drug using a database of curated gene-drug relationships and the collective gene impact) to identify the top 25 therapeutic candidates in the PT samples. Afterward, we performed a new query with PanDrugs using the exonic mutations in each CTC sample to examine whether the CTC-derived genomic information identifies similar therapeutic options and drug sensitivity scores as for the bulk.

2.9. MATH scores

For the datasets with an unknown number of CTCs or with CTC counts >1, we additionally estimated the mutant-allele tumor heterogeneity (MATH) score [31]. The MATH score is based on the distribution of allele fractions among somatic mutations. It is calculated as the percentage ratio of the width of the data to the center of its distribution:

$$MATH = 100 \times \frac{median\ absolute\ deviation\ (VAF)}{median\ (VAF)}$$

Importantly, since MATH scores are sensitive to unreliable allele frequency estimates stemming from sites with poor sequencing coverage depth, for their calculation, we applied an additional filter to restrict our mutation calls to positions showing a depth of coverage ≥25 and alternative allelic depth ≥ 5.

3. Results

3.1. CTC-counts

Patient-level CTC counts were available for two of the selected CTC-capture methods (as the Parsortix platform does not provide CTC counts). Using the CellSearch system, we isolated one CTC from patient P1, whereas we did not detect CTCs for the remaining patients. Importantly, although we were able to recover potential CTCs from all patients using FACS (P1 = 2 CTCs; P3 = 6 CTCs; P4 = 2 CTCs; P5 = 1 CTC), after whole-genome amplification we only obtained high-quality sequencing libraries for patients P4 and P5.

3.2. Tumor mutational burden

Across all patients, we found sharp differences in the number of somatic mutations (SNVs) identified with the different CTC-capture methods (Fig. 1). While the number of mutations called with both CellSearch and Parsortix datasets was close to, or lower than, the number of mutations observed in the matched PT bulk samples, the mutation counts in the FACS CTC pools were consistently much higher (by one order of magnitude) than in the bulk samples. Remarkably, regardless of the CTC-capture strategy, we found a minimal overlap in mutation calls between CTC pools and PT datasets: Parsortix (average of 4.6%), CellSearch (3.4%), and FACS (0.4%).

Similarly, within each patient, the number of mutations shared between CTC pools captured with distinct methods was generally small (Fig. 2a). We identified many CTC-specific mutations across datasets, including non-silent ones (Fig. 2a and Fig. S4).

Furthermore, within the PT samples, the median variant allele frequency (VAF) of the shared mutations with CTC pools ranged from 0.19 (Parsortix - P4) to 0.09 (FACS - P5), suggesting that the isolated CTCs derive from minor subclones within the PTs (Fig. 2b). We observed similar VAF scores between the CellSearch and Parsortix CTCs. In P4, we found a significant difference between Parsortix and FACS (Fig. 2b), with the FACS CTCs being generally enriched with mutations at low frequency.

As shown in Fig. 2b, the CTC pools showed an evident depletion of clonal (i.e., allele frequency ≥ 0.4) mutations observed in the PT samples. While these results may appear surprising, as, in theory, clonal mutations in the PT should appear in all CTCs sampled, the failure to identify such mutations can be partly explained due to the limited coverage breadth of the CTCs. Indeed, across all samples, only a small

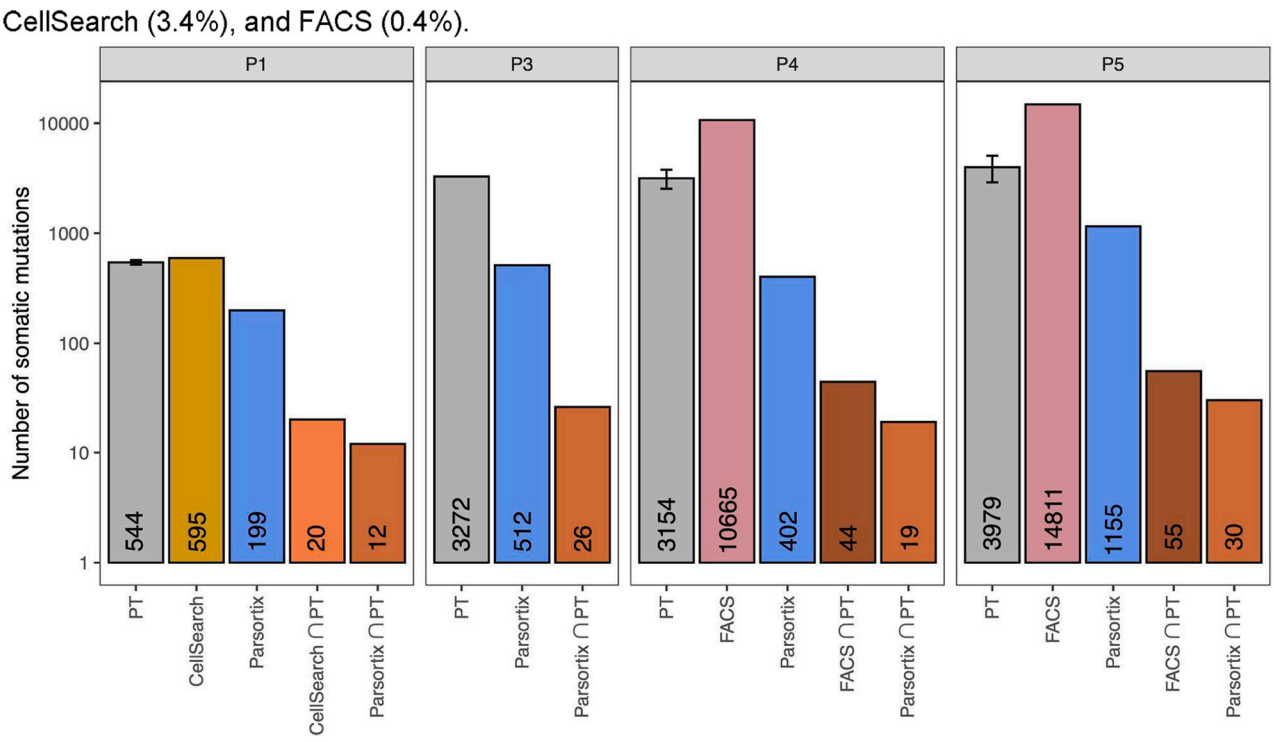

Fig. 1. SNV abundance per CTC-capturing method. Barplots depicting the total number of SNVs identified by MuTect2 for each dataset. The number of SNVs is shown at the bottom of each bar. Bar colors reflect the different input material or capturing method: gray = FFPE primary tumor (PT) sample; gold = CellSearch CTC pool; blue = Parsortix CTC pool; pink = FACS CTC pool. Error bars are only available for bulk tumor samples and reflect the 95% confidence interval. Orange bars depict the number of shared sites between CTC datasets and PT samples. The y-axis is on the log scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

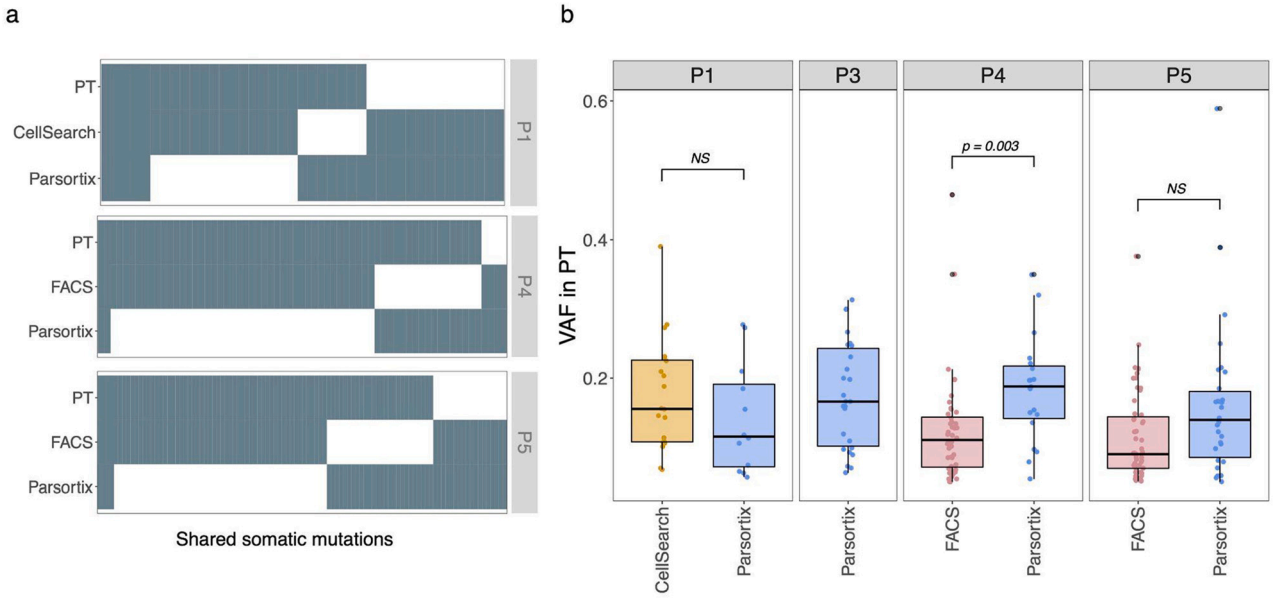

Fig. 2. Genomic profiling and clonality of CTC-pools. a. Occupancy matrix of shared sites (P1 = 41 SNVs; P4 = 65 SNVs; P5 = 100 SNVs) across the CTC-capture methods within patients. Different colored tiles reflect different mutation status: dark gray = mutation; white = reference/missing data. Patient ID is shown on the right. b. Boxplots depicting the bulk-level VAF estimates of the SNVs shared with the CTC pools. Boxplot colors represent the different CTC-capture methods. Statistical analysis was performed using the two-sample KS-test to compare bulk VAF estimates of shared sites between CTC-capture methods. Significant *p-values* are shown above boxplots.

fraction (5 to 39%) of regions harboring clonal variants in the PTs were covered with sequencing reads in the CTC pools (Table SI). Moreover, after looking at the coverage statistics of heterozygous single-nucleotide polymorphisms (SNPs) (Fig. S5), we additionally found strong evidence of allele dropout (ADO) taking place during WGA. Despite being particularly obvious in the FACS datasets -which showed an averaged ADO of 67% for the heterozygous sites called-, all CTC pools showed some degree of ADO (CellSearch ADO: 9% and Parsortix ADO: 9%)

which very possibly interfered with the identification of clonal mutations.

3.3. Measuring intratumor genomic heterogeneity (ITH)

Afterward, we explored potential differences in intratumor genomic heterogeneity (ITH) estimates between PT and CTC samples by examining the frequency distribution of somatic mutations across the different datasets. Significantly, since several of our CTC samples comprised only one cell, the subsequent analyses were limited to datasets with either an unknown number of CTCs or CTC counts >1 and included the Parsortix CTCs from patients P1, P3, P4, and P5, and the FACS CTCs from patient P4.

As illustrated in Fig. 3a-b (and Fig. S6a-b), we found significant differences in the distribution of VAF estimates between PT and CTC pools. In patient P4, both FACS and Parsortix datasets showed an enrichment towards low-frequency variants compared to the corresponding PT sample. Moreover, a quantile-quantile (Q-Q) plot further revealed contrasting differences in the skewness of the VAF distributions stemming from the different CTC-capture methods (Fig. 3b). Indeed, while the Parsortix CTCs encompassed a more significant number of mutations at intermediate frequencies when compared to the matched PT sample, the FACS CTCs showed a substantial depletion of this class of mutations.

Interestingly, in the Parsortix CTCs, we found a significant positive correlation between the VAF estimates of shared mutations between the CTC pool and the PT sample (Fig. 3c). As for the FACS CTCs, although we found a positive trend, the relationship was not significant (similar to the results of the remaining datasets - Fig. S6c).

We also observed sharp differences in the ITH levels, measured by MATH scores, among CTC-capture methods. In patient P4, the Parsortix CTCs and the PT sample returned highly concordant MATH scores (43.6 and 43.0, respectively), while the FACS dataset displayed a much larger MATH score (62.8) (Fig. 3d). We observed similar MATH scores between the remaining Parsortix CTCs and the PT samples (Fig. S6d).

3.4. Mutational signatures

We next explored the mutational signatures in the different datasets to look for potential differences between CTC pools and PT samples (Fig. 4a). Across all datasets, *SigProfilerExtractor* identified five mutational processes, with PT samples predominantly enriched in "clock-like" (i.e., aging) COSMIC signatures SBS1 and SBS5. A similar contribution of SBS1 and SBS5 was also found in the CellsSearch and Parsortix CTC datasets, albeit the lower cosine similarity score (cosine similarity: 0.505) obtained - likely due to the limited number of mutations available for signature assignment in these datasets. In contrast, FACS CTCs were predominantly characterized by signature SBS46 (65% in both cases), a mutational signature associated with sequencing artifacts.

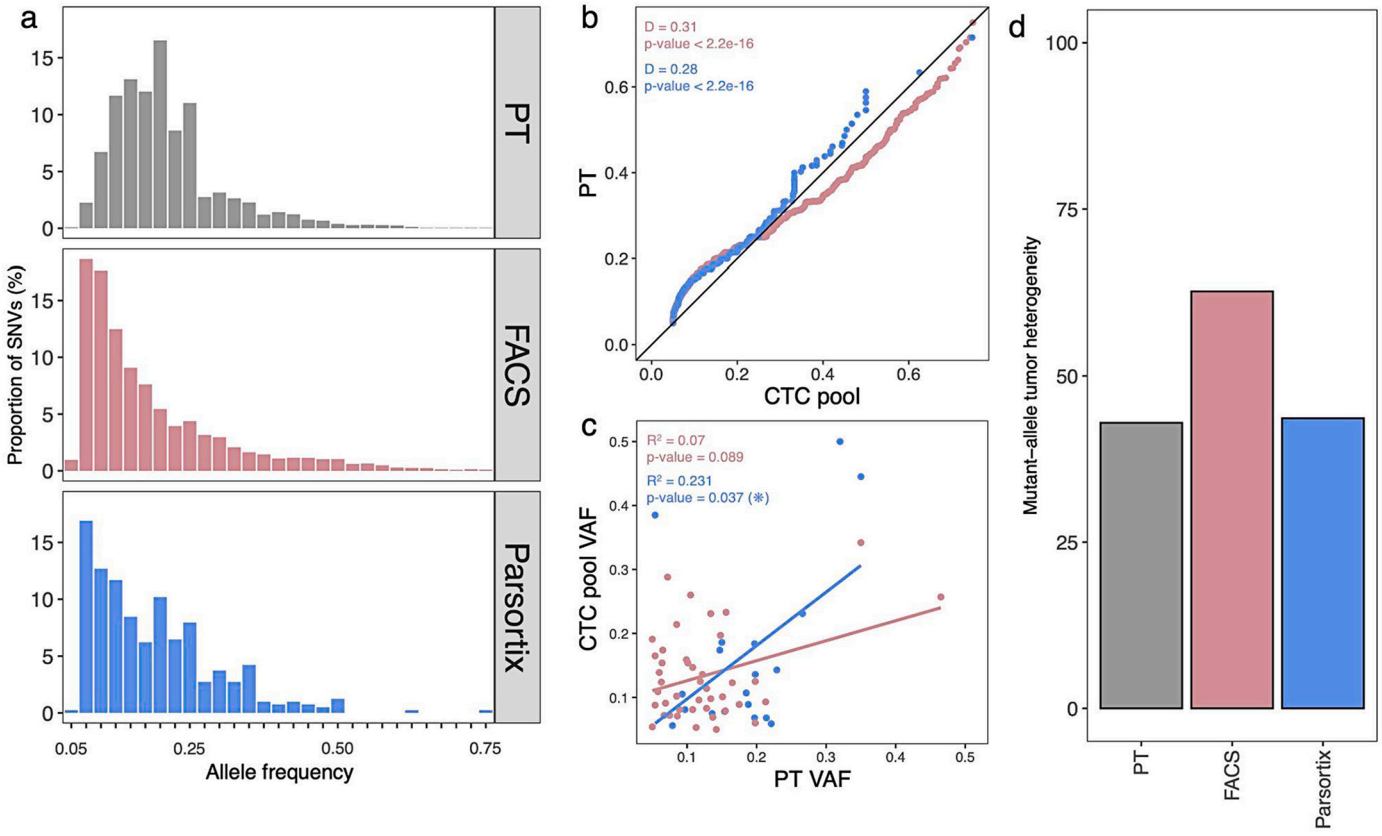

Fig. 3. Measuring ITH through CTC pools. a. Histograms depicting the variant allele frequency (VAF) distribution of somatic mutations for the different datasets of patient P4. Different datasets highlighted with different colors: PT = gray; FACS = pink; Parsortix = blue with dataset ID shown on the right. Histograms are scaled to percentages. b. Q-Q plot compares allele frequency estimates in CTC pools and PT samples of patient P4. Statistical analysis was performed using the two-sample KS-test to compare the VAF distribution between CTC pools and PT samples. KS D (distance) statistic and *p-values* are shown on the upper left side of the plot. c. Scatter plot describing the similarity of VAF scores of overlapping sites between CTC pools and PT samples. Solid lines represent the best fit from regression analysis. R^2 scores and *p-values* are shown on the upper left side of the plot. d. Barplot depicting the MATH scores obtained using the mutation sets passing our strict filtering - Primary tumor = 416 SNVs; FACS = 8193 SNVs; Parsortix = 46 SNVs (see *methods*). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

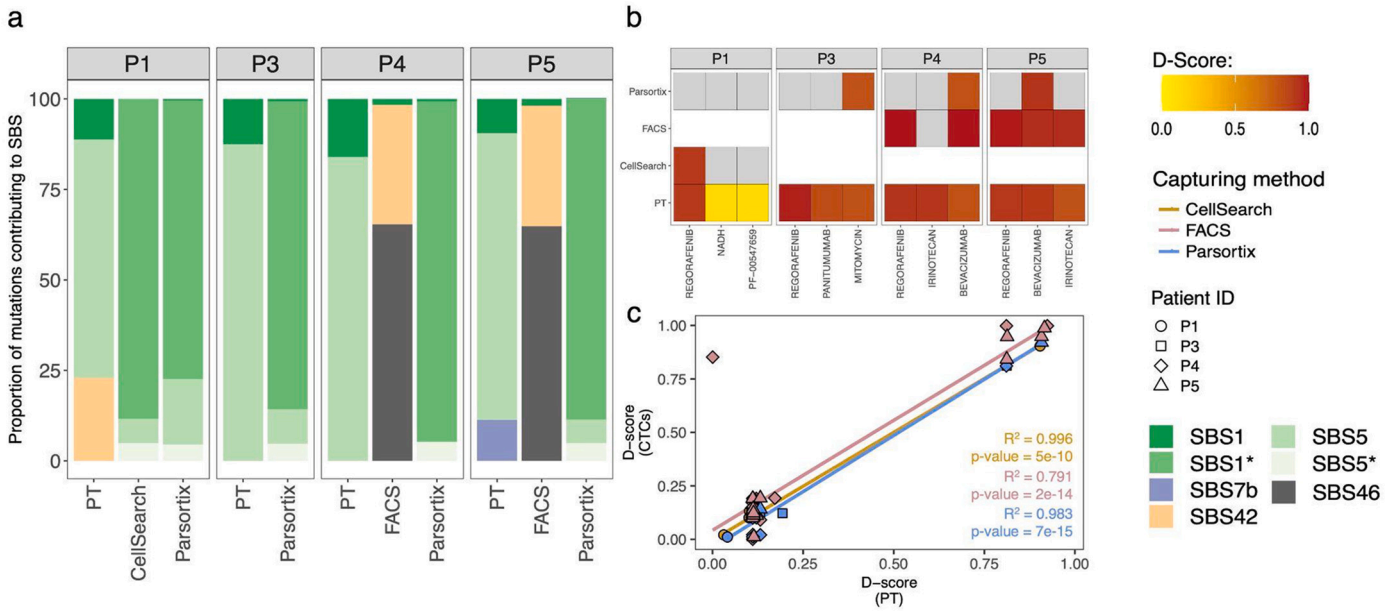

Fig. 4. Mutational signatures and drug sensitivity concordance. a. Barplots depict the proportion of mutations contributing to the different signatures/processes across the samples. Sample IDs are shown at the bottom, and patient IDs are at the top. Different colors reflect the identified mutational signature with the COSMIC SBS ID displayed at the right of the plot. Legend asterisks distinguish mutational signatures identified by *SigProfilerExtractor* showing suboptimal cosine similarity scores (i.e., < 0.90). b. Tile plots depicting the overlap between the top-3 drug candidates in the primary tumor (PT) samples (ordered from left to right) and corresponding CTC-pools. Tiles are colored according to the drug sensitivity score (D-score). Gray tiles correspond to therapeutic options not recovered by *PanDrugs* for that specific sample. c. Scatter plot depicting the correlation between the best therapeutic candidates (measured using the D-score) identified in PT and CTC samples. Shape distinguishes the different patients, while colors reflect the different CTC-capture methods. Solid lines represent the best fit from regression analysis. R^2 scores and *p-values* are shown on the bottom right side of the plot.

3.5. Drug suitability scores

Finally, using the list of exonic positions available, we collected the best drug candidates for each dataset to evaluate potential changes in the type and sensitivity of therapeutic options between CTC pools and PT samples (Fig. 4b). Across patients, both FACS and CellSearch CTCs recovered, in most cases, the top candidate drugs (i.e., with the highest D-Score) identified with the corresponding PT samples. The Parsortix CTCs, on the other hand, often returned only a subset of treatment options specified for the matched PT. All CTC-capture methods showed a significant positive correlation in drug sensitivity scores between CTCs and PT samples (Fig. 4c,) suggesting that the CTC genomic profiles offer reliable information for prioritizing therapeutic strategies in CRC.

4. Discussion

The isolation of CTCs remains challenging mainly because of their scarcity in peripheral blood and phenotypic heterogeneity. While recent technological advances have improved capture strategies, most available methods differ in crucial aspects such as enrichment efficiency, cell viability, and throughput. Studies evaluating the implications of different CTC-capture methods on downstream analyses are scant, and it remains unclear whether distinct enrichment approaches can provide compatible descriptions of the mutational landscape of CTCs.

In this study, to evaluate the impact of different CTC-capture strategies on the downstream molecular characterization of CTCs, we contrasted the genomic profiles of primary tumor samples against CTC pools recovered using three different enrichment strategies in four mCRC patients.

Our whole-exome sequencing experiments suggest differences in the mutational loads of CTC pools due to the enrichment method used. In sharp contrast to the results obtained with the CellSearch and Parsortix systems, the mutation counts in our FACS CTC pools exceeded the number of mutations observed in the corresponding primary tumor samples by an order of magnitude. Notably, the FACS datasets showed significant enrichment for mutations at lower frequencies. These mutations were later linked to a mutational signature associated with sequencing artifacts (i.e., SBS46).

These results suggest that our FACS-derived CTCs might have accumulated DNA lesions along with the different steps of the protocol. Indeed, FACS-sorted CTCs are fixed with a solution containing formaldehyde, which might induce some degree of DNA damage. However, since formaldehyde is also present in the collection tubes used with the Parsortix system, it is unlikely that this is the main reason for the large number of presumably spurious mutations seen in the FACS-sorted CTCs. Unfortunately, identifying and subsequently removing these potential errors is not necessarily straightforward. Although one could argue that setting lower and upper bounds on the minor allele frequency could potentially prevent downstream variant call artifacts, we should note that multiple rounds of genomic amplification must precede sequencing CTC pools. Since most WGA methods inevitably introduce biases in the resulting sequencing data (e.g., uneven genome amplification, allelic imbalances, and dropout), for any given site, the derived allele frequency score will not necessarily reflect its actual frequency in the CTC population sampled. Indeed, our datasets reflected these biases in the imbalanced distribution of allele frequencies of heterozygous SNPs and the relatively poor concordance in VAFs of shared somatic sites between CTC pools and PT samples.

In any case, and for all datasets analyzed, Parsortix-derived CTC pools provided similar descriptions of ITH when compared to the corresponding PT samples, with MATH scores in explicit agreement with previous estimates in mCRC [32]. In contrast, this metric was vastly overestimated in our FACS dataset, perhaps due to many potential sequencing artifacts. Moreover, while the identification of therapeutic candidates was not always identical between CTC pools and PT samples, all CTC pools analyzed suggested drug candidates displaying a significant probability of response and highly concordant sensitivity scores with the PT samples, thus providing strong evidence that CTC-based

mutational profiles may contribute with valuable guidance for refining treatment tailoring [33–35].

5. Conclusion

In conclusion, CTC genomics still faces technical challenges for straightforward clinical applications. As seen throughout our study, all methods evaluated struggled with data quality issues –potentially caused by the inherent technical bias introduced by the limiting amounts of input material [36] and background DNA contamination (e.g., white blood cells) [37] – which resulted in a somewhat incomplete picture of the mutational landscape of these tumors. Moreover, the limited number of patients analyzed and the failure of some enrichment methods to recover CTCs for all patients, may prevent the findings from this study from being fully generalizable. Still, it is essential to highlight that the CTC pools recovered from Parsortix and, in part, CellSearch returned comparable ITH estimates, similar mutational signature profiles, and suggested equivalent therapeutic targets, when compared to those found in matching primary tumor samples.

On this basis, as the performance of technologies for the detection and isolation of CTCs continues to improve, allowing for more accurate and informative genomic data to be produced [38], future studies using larger cohorts should explore whether the mutational landscape and genomic diversity of CTC populations can indeed provide clinically relevant prognostic and predictive information beyond simple enumeration.

Ethics approval and consent to participate

All specimens used in this study were obtained and collected after written informed consent from all subjects using a protocol approved by the Clinical Ethics Committee of Pontevedra-Vigo-Ourense (2018/301 approved 19/06/2018).

Availability of data and materials

We have deposited raw whole-exome sequencing data at the Sequence Read Archive database under the accession code PRJNA886718.

Funding

This work was supported by an AXA Research Fund postdoctoral grant (awarded to J.M.A) and the Spanish Ministry of Science and Innovation - MICINN (PID2019-106247GB-I00 awarded to D.P.). J.M.A. is currently supported by the AECC (INVES20007FERN). D.P. receives further support from Xunta de Galicia. M.V. is supported by an H2020/Marie Skłodowska-Curie Actions EU research framework programme grant (Project H2020 MSCA-ITN-2017-766030). L.T. received a Ph.D. fellowship from Xunta de Galicia (ED481A-2018/303). J.C. received grants from Spain's Carlos III Health Care Institute (Co-funded by European Regional Development Fund/European Social Fund "A way to make Europe"/"Investing in your future"), No. PI17/00837 and PI21/01771. J.C. is additionally funded by the Axencia Galega de Innovación (N607B-2020/02). R.P. received support from Roche-Chus Joint Unit (IN853B 2018/03) funded by Axencia Galega de Innovación (GAIN), Consellería de Economía, Emprego e Industria.

Authors' contributions.

D.P. and J.M.A. conceived the study and designed the analyses. M.S. and J.C. collected the blood and tumor samples and patient information. N.E.G., S.P.L., and P.A. processed the samples, performed the spike-in experiments, obtained the CTC pools through FACS, and performed whole-genome amplification. R.P, L.M.R, and P.M collected the CTC pools from CellSearch and Parsortix. J.M.A., L.T., and M.V. performed the analyses. All authors read and approved the final manuscript.

Declaration of Competing Interest

The authors declare no competing interests.

Data availability

We have deposited raw whole-exome sequencing data at the Sequence Read Archive database under the accession code PRJNA886718.

Acknowledgments

We thank all lab members for their comments on earlier manuscript versions. We also thank the Supercomputation Center of Galicia (CESGA) for providing all computational resources and Mercedes Peleteiro, the flow cytometry core facility manager, for her help and support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2022.110500.

References

[1] J. Marrugo-Ramírez, M. Mir, J. Samitier, Blood-based Cancer biomarkers in liquid biopsy: a promising non-invasive alternative to tissue biopsy, Int. J. Mol. Sci. 19 (2018), https://doi.org/10.3390/ijms19102877.
[2] E.G. Robertson, G. Baxter, Tumour seeding following percutaneous needle biopsy: the real story!, Clin. Radiol. 66 (2011) 1007–1014.
[3] K.R. Chi, The tumour trail left in blood, Nature. 532 (2016) 269–271, https://doi.org/10.1038/532269a.
[4] K. Pantel, M.R. Speicher, The biology of circulating tumor cells, Oncogene. 35 (2016) 1216–1224, https://doi.org/10.1038/onc.2015.192.
[5] E. Crowley, F. Di Nicolantonio, F. Loupakis, A. Bardelli, Liquid biopsy: monitoring cancer-genetics in the blood, Nat. Rev. Clin. Oncol. 10 (2013) 472–484.
[6] M.G. Krebs, R.L. Metcalf, L. Carter, G. Brady, F.H. Blackhall, C. Dive, Molecular analysis of circulating tumour cells-biology and biomarkers, Nat. Rev. Clin. Oncol. 11 (2014) 129–144.
[7] K. Pantel, C. Alix-Panabières, Circulating tumour cells and cell-free DNA in gastrointestinal cancer, nature reviews, Gastroenterol. Hepatol. 14 (2017) 73–74, https://doi.org/10.1038/nrgastro.2016.198.
[8] F. Castro-Giner, N. Aceto, Tracking cancer progression: from circulating tumor cells to metastasis, Genome Med. 12 (2020) 31.
[9] M.J.M. Magbanua, C. Yau, D.M. Wolf, J.S. Lee, A. Chattopadhyay, J.H. Scott, E. Bowlby-Yoder, E. Shelley Hwang, M. Alvarado, C.A. Ewing, A.L. Delson, L. J. van't Veer, L. Esserman, J.W. Park, Synchronous detection of circulating tumor cells in blood and disseminated tumor cells in bone marrow predicts adverse outcome in early breast cancer, Clin. Cancer Res. 25 (2019) 5388–5397, https://doi.org/10.1158/1078-0432.ccr-18-3888.
[10] A.B. Silveira, F.-C. Bidard, M.-L. Tanguy, E. Girard, O. Trédan, C. Dubot, W. Jacot, A. Goncalves, M. Debled, C. Levy, J.-M. Ferrero, C. Jouannaud, M. Rios, M.-A. Mouret-Reynier, F. Dalenc, C. Hego, A. Rampanou, B. Albaud, S. Baulande, F. Berger, J. Lemonnier, S. Renault, I. Desmoulins, C. Proudhon, J.-Y. Pierga, Multimodal liquid biopsy for early monitoring and outcome prediction of chemotherapy in metastatic breast cancer, Npj Breast Cancer. 7 (2021), https://doi.org/10.1038/s41523-021-00319-4.
[11] U. Basso, A. Facchinetti, E. Rossi, M. Maruzzo, V. Conteduca, M. Aieta, F. Massari, A.P. Fraccon, C. Mucciarini, T. Sava, M. Santoni, C. Pegoraro, E. Durante, M. Nicodemo, A. Perin, A. Bearz, C. Gatti, P. Fiduccia, A. Diminutto, C. Barile, U. De Giorgi, R. Zamarchi, V. Zagonel, Prognostic role of circulating tumor cells in metastatic renal cell carcinoma: a large, multicenter, prospective trial, Oncologist 26 (2021) 740–750, https://doi.org/10.1002/onco.13842.
[12] F. Chemi, D.G. Rothwell, N. McGranahan, S. Gulati, C. Abbosh, S.P. Pearce, C. Zhou, G.A. Wilson, M. Jamal-Hanjani, N. Birkbak, J. Pierce, C.S. Kim, S. Ferdous, D.J. Burt, D. Slane-Tan, F. Gomes, D. Moore, R. Shah, M. Al Bakir, C. Hiley, S. Veeriah, Y. Summers, P. Crosbie, S. Ward, B. Mesquita, M. Dynowski, D. Biswas, J. Tugwood, F. Blackhall, C. Miller, A. Hackshaw, G. Brady, C. Swanton, C. Dive, TRACERx consortium, pulmonary venous circulating tumor cell dissemination before tumor resection and disease relapse, Nat. Med. 25 (2019) 1534–1539.
[13] N. Gulbahce, M.J.M. Magbanua, R. Chin, M.R. Agarwal, X. Luo, J. Liu, D. M. Hayden, Q. Mao, S. Ciotlos, Z. Li, Y. Chen, X. Chen, Y. Li, R.Y. Zhang, K. Lee, R. Tearle, E. Park, S. Drmanac, H.S. Rugo, J.W. Park, R. Drmanac, B.A. Peters, Quantitative whole genome sequencing of circulating tumor cells enables personalized combination therapy of metastatic Cancer, Cancer Res. 77 (2017) 4530–4541.
[14] M.J.M. Magbanua, H.S. Rugo, D.M. Wolf, L. Hauranieh, R. Roy, P. Pendyala, E. V. Sosa, J.H. Scott, J.S. Lee, B. Pitcher, T. Hyslop, W.T. Barry, S.J. Isakoff, M. Dickler, L. Van't Veer, J.W. Park, Expanded genomic profiling of circulating tumor cells in metastatic breast cancer patients to assess biomarker status and

biology over time (CALGB 40502 and CALGB 40503, Alliance), Clin. Cancer Res. 24 (2018) 1486–1499.
[15] C.M. Court, S. Hou, L. Liu, P. Winograd, B.J. DiPardo, S.X. Liu, P.-J. Chen, Y. Zhu, M. Smalley, R. Zhang, S. Sadeghi, R.S. Finn, F.M. Kaldas, R.W. Busuttil, X.J. Zhou, H.-R. Tseng, J.S. Tomlinson, T.G. Graeber, V.G. Agopian, Somatic copy number profiling from hepatocellular carcinoma circulating tumor cells, NPJ Precis Oncol. 4 (2020) 16.
[16] M. Tellez-Gabriel, B. Ory, F. Lamoureux, M.-F. Heymann, D. Heymann, Tumour heterogeneity: the key advantages of single-cell analysis, Int. J. Mol. Sci. 17 (2016), https://doi.org/10.3390/ijms17122142.
[17] E. Pailler, V. Faugeroux, M. Oulhen, L. Mezquita, M. Laporte, A. Honoré, Y. Lecluse, P. Queffelec, M. NgoCamus, C. Nicotra, J. Remon, L. Lacroix, D. Planchard, L. Friboulet, B. Besse, F. Farace, Acquired resistance mutations to ALK inhibitors identified by single circulating tumor cell sequencing in -rearranged non-small-cell lung Cancer, Clin. Cancer Res. 25 (2019) 6671–6682.
[18] C. Paoletti, A.K. Cani, J.M. Larios, D.H. Hovelson, K. Aung, E.P. Darga, E. M. Cannell, P.J. Baratta, C.-J. Liu, D. Chu, M. Yazdani, A.R. Blevins, V. Sero, N. Tokudome, D.G. Thomas, C. Gersch, A.F. Schott, Y.-M. Wu, R. Lonigro, D. R. Robinson, A.M. Chinnaiyan, F.Z. Bischoff, M.D. Johnson, B.H. Park, D.F. Hayes, J.M. Rae, S.A. Tomlins, Comprehensive mutation and copy number profiling in archived circulating breast Cancer tumor cells documents heterogeneous resistance mechanisms, Cancer Res. 78 (2018) 1110–1122.
[19] E. Rossi, R. Zamarchi, Single-cell analysis of circulating tumor cells: how far have we come in the -omics era? Front. Genet. 0 (2019) https://doi.org/10.3389/fgene.2019.00958.
[20] A. Kowalik, M. Kowalewska, S. Góźdź, Current approaches for avoiding the limitations of circulating tumor cells detection methods—implications for diagnosis and treatment of patients with solid tumors, Transl. Res. 185 (2017) 58–84.e15, https://doi.org/10.1016/j.trsl.2017.04.002.
[21] P. Bankó, S.Y. Lee, V. Nagygyörgy, M. Zrínyi, C.H. Chae, D.H. Cho, A. Telekes, Technologies for circulating tumor cell separation from whole blood, J. Hematol. Oncol. 12 (2019), https://doi.org/10.1186/s13045-019-0735-4.
[22] R.P.L. Neves, W. Ammerlaan, K.C. Andree, S. Bender, L. Cayrefourcq, C. Driemel, C. Koch, M.V. Luetke-Eversloh, M. Oulhen, E. Rossi, C. Alix-Panabières, F. Betsou, F. Farace, S. Riethdorf, T. Schlange, H. Wikman, R. Zamarchi, K. Pantel, L.W.M. M. Terstappen, N.H. Stoecklein, CANCER-ID consortium, proficiency testing to assess technical performance for CTC-processing and detection methods in CANCER-ID, Clin. Chem. 67 (2021) 631–641.
[23] K.-A. Hyun, G.-B. Koo, H. Han, J. Sohn, W. Choi, S.-I. Kim, H.-I. Jung, Y.-S. Kim, Epithelial-to-mesenchymal transition leads to loss of EpCAM and different physical properties in circulating tumor cells from metastatic breast cancer, Oncotarget. 7 (2016) 24677–24687.
[24] Cell enrichment from human blood through red cell lysis, in: Methods in Cell Biology, Academic Press, 2012, pp. 183–192.
[25] H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, arXiv [q-bio.GN] (2013). http://arxiv.org/abs/1303.3997.
[26] G.A. Van der Auwera, M.O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K.V. Garimella, D. Altshuler, S. Gabriel, M.A. DePristo, From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline, Curr. Protoc. Bioinformatics 43 (2013) 11.10.1–11.10.33.
[27] K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data, Nucleic Acids Res. 38 (2010), e164.
[28] S.M. Ashiqul Islam, Y. Wu, M. Díaz-Gay, E.N. Bergstrom, Y. He, M. Barnes, M. Vella, J. Wang, J.W. Teague, P. Clapham, S. Moody, S. Senkin, Y.R. Li, L. Riva, T. Zhang, A.J. Gruber, R. Vangara, C.D. Steele, B. Otlu, A. Khandekar, A. Abbasi, L. Humphreys, N. Syulyukina, S.W. Brady, B.S. Alexandrov, N. Pillay, J. Zhang, D. J. Adams, I. Martincorena, D.C. Wedge, M.T. Landi, P. Brennan, M.R. Stratton, S. G. Rozen, L.B. Alexandrov, Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor, bioRxiv (2021), https://doi.org/10.1101/2020.12.13.422570.
[29] L.B. Alexandrov, J. Kim, N.J. Haradhvala, M.N. Huang, A.W. Tian Ng, Y. Wu, A. Boot, K.R. Covington, D.A. Gordenin, E.N. Bergstrom, S.M.A. Islam, N. Lopez-Bigas, L.J. Klimczak, J.R. McPherson, S. Morganella, R. Sabarinathan, D. A. Wheeler, V. Mustonen, G. Getz, S.G. Rozen, M.R. Stratton, The repertoire of mutational signatures in human cancer, Nature. 578 (2020) 94–101.
[30] E. Piñeiro-Yáñez, M. Reboiro-Jato, G. Gómez-López, J. Perales-Patón, K. Troulé, J. M. Rodríguez, H. Tejero, T. Shimamura, P.P. López-Casas, J. Carretero, A. Valencia, M. Hidalgo, D. Glez-Peña, F. Al-Shahrour, PanDrugs: a novel method to prioritize anticancer drug treatments according to individual genomic data, Genome Med. 10 (2018) 41.
[31] E.A. Mroz, J.W. Rocco, MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma, Oral Oncol. 49 (2013) 211–215, https://doi.org/10.1016/j.oraloncology.2012.09.007.
[32] F. Bettoni, C. Masotti, B.R. Corrêa, E. Donnard, F.F. dos Santos, G.P. São Julião, B. B. Vailati, A. Habr-Gama, P.A.F. Galante, R.O. Perez, A.A. Camargo, The effects of neoadjuvant Chemoradiation in locally advanced rectal cancer—the impact in Intratumoral heterogeneity, Front. Oncol. 9 (2019), https://doi.org/10.3389/fonc.2019.00974.
[33] B.L. Khoo, G. Grenci, T. Jing, Y.B. Lim, S.C. Lee, J.P. Thiery, J. Han, C.T. Lim, Liquid biopsy and therapeutic response: circulating tumor cell cultures for evaluation of anticancer treatment, Sci. Adv. 2 (2016), e1600274.
[34] G. Siravegna, S. Marsoni, S. Siena, A. Bardelli, Integrating liquid biopsies into the management of cancer, nature reviews, Clin. Oncol. 14 (2017) 531–548, https://doi.org/10.1038/nrclinonc.2017.14.
[35] A.R. Parikh, I. Leshchiner, L. Elagina, L. Goyal, C. Levovitz, G. Siravegna, D. Livitz, K. Rhrissorrakrai, E.E. Martin, E.E. Van Seventer, M. Hanna, K. Slowik, F. Utro, C. J. Pinto, A. Wong, B.P. Danysh, F.F. de la Cruz, I.J. Fetter, B. Nadres, H. A. Shahzade, J.N. Allen, L.S. Blaszkowsky, J.W. Clark, B. Giantonio, J.E. Murphy, R.D. Nipp, E. Roeland, D.P. Ryan, C.D. Weekes, E.L. Kwak, J.E. Faris, J.Y. Wo, F. Aguet, I. Dey-Guha, M. Hazar-Rethinam, D. Dias-Santagata, D.T. Ting, A.X. Zhu, T.S. Hong, T.R. Golub, A.J. Iafrate, V.A. Adalsteinsson, A. Bardelli, L. Parida, D. Juric, G. Getz, R.B. Corcoran, Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers, Nat. Med. 25 (2019) 1415–1421.
[36] N.E. Navin, Cancer genomics: one cell at a time, Genome Biol. 15 (2014), https://doi.org/10.1186/s13059-014-0452-9.
[37] L. Xu, X. Mao, A. Imrali, F. Syed, K. Mutsvangwa, D. Berney, P. Cathcart, J. Hines, J. Shamash, Y.-J. Lu, Optimization and evaluation of a novel size based circulating tumor cell isolation system, PLoS One 10 (2015), e0138032.
[38] Z. Diamantopoulou, F. Castro-Giner, N. Aceto, Circulating tumor cells: ready for translation? J. Exp. Med. 217 (2020) https://doi.org/10.1084/jem.20200356.